Statistique et analyse des données

Alain Degenne
Claude Flament
Pierre Verges

Approche galoisienne en classification

Statistique et analyse des données, tome 3, n° 1 (1978), p. 29-45
<http://www.numdam.org/item?id=SAD_1978__3_1_29_0>

Statistique et Analyse des Données
1 - 1978

APPROCHE GALOISIENNE EN CLASSIFICATION

par Alain DEGENNE (1), Claude FLAMENT (2) et Pierre VERGES (3)

Classe : ensemble d'objets défini par le fait que ces objets possèdent tous et possèdent seuls un ou plusieurs caractères communs.

LALANDE, Vocabulaire de la Philosophie, 1926

En s'inspirant de cette définition, on peut proposer une démarche classificatoire radicalement différente des méthodes habituelles qui s'appuient sur un indice numérique de ressemblance entre les objets : en effet, la formalisation de la définition de Lalande se fait normalement en utilisant les *correspondances de Galois*.

CORRESPONDANCE DE GALOIS ASSOCIEE A UN TABLEAU DE CORRESPONDANCE (4)

Soit (X,Y) un tableau de correspondance "en 0,1", c'est-à-dire le tableau d'une rela-

---

(1) Laboratoire d'Economie et de Sociologie du Travail, Aix-en-Provence.

(2) Département de Psychologie, Université de Provence, Aix-en-Provence.

(3) Centre de Mathématique Sociale, Ecole des Hautes Etudes en Sciences Sociales, Marseille.

(4) Cf. M. BARBUT et B. MONJARDET, *Ordre et classification : Algèbre et combinatoire*, Paris, Hachette, 1970 (Tome 2, pp. 7 à 33).

tion ρ de X vers Y. On sait depuis BIRKHOFF (*Lattice theory*, 1940), associer à un tel tableau une correspondance de Galois (α,β) entre les ensembles $\mathcal{P}(X)$ et $\mathcal{P}(Y)$ des parties de X et de Y :

- pour toute partie non vide A de X, on définit αA comme la partie de Y dont chaque élément y correspond à chaque élément x de A.

$$\alpha A = \{y : y \in Y, x \rho y \ \forall x \in A\} ;$$

si A est la partie vide, on pose : $\alpha A = Y$.

- dualement, pour toute partie non vide B de Y, on a :

$$\beta B = \{x : x \in X, x \rho y \quad \forall y \in B\} ;$$

si B est la partie vide, on pose $\beta B = X$.

Rappelons les principales propriétés :

- les applications α et β sont antitones :

$$A \subset A' \subset X \Rightarrow \alpha A \supset \alpha A', \text{ et } B \subset B' \subset Y \Rightarrow \beta B \supset \beta B' ;$$

- l'application composée $\beta\alpha$ de $\mathcal{P}(X)$ dans lui-même, est une fermeture, c'est-à-dire est extensive ($A \subset \beta\alpha A$, $A \subset X$), isotone ($A \subset A' \subset X \Longrightarrow \beta\alpha A \subset \beta\alpha A'$) et idempotente ($\beta\alpha\beta\alpha A = \beta\alpha A$, $A \subset X$) ;

- dualement, $\alpha\beta$ est une fermeture dans $\mathcal{P}(Y)$.

INTERPRETATION

Si X est un ensemble d'objets et Y un ensemble de caractères, on interprète la relation ρ comme signifiant la possession d'un caractère y par un objet x : $x\rho y$. On voit alors que toute partie de X, fermée pour $\beta\alpha$, est une classe au sens de Lalande, et réciproquement ; en effet, soit une partie fermée C de X : $\beta\alpha C = C$; les éléments de C possèdent en commun exactement les caractères constituant l'ensemble αC, et ils sont les seuls à posséder tous ces caractères. Supposons en effet qu'un élément x de X possède ces caractères. On a : $\alpha C \subset \alpha x$; d'où, par l'antitonie de β : $\beta\alpha x \subset \beta\alpha C$; et, puisque $x \in \beta\alpha x$ (extensivité) et que $\beta\alpha C = C$ (par hypothèse), on a : $x \in C$.

DEFINITION

On dira qu'une partition $\mathcal{G} = (G_i)$ de X est une *partition galoisienne* sur le tableau (X,Y), si et seulement si chaque classe G_i de $\mathcal{G}$ est *fermée* pour la fermeture galoisienne associée au tableau :

$$G_i = \beta\alpha G_i, \ \forall i.$$

Notons que la partition triviale réduite à une seule classe est galoisienne, puisque $X \subset \beta\alpha X \subset X$. Cette partition galoisienne n'a évidemment aucun intérêt autre que de faciliter certaines démonstrations.

PROPRIETE

Soit (X,Y) un tableau de correspondance et une partition quelconque $\mathcal{C} = (C_i)$ de X. L'ensemble des partitions galoisiennes moins fines que $\mathcal{C}$ admet un plus petit élément appelé classification galoisienne associée à $\mathcal{C}$.

Corollaire : Pour tout tableau de correspondance, il existe une classification galoisienne ξ_o plus fine que toutes les autres.

En effet, l'ensemble des partitions galoisiennes plus grossières que $\mathcal{C}$ n'est pas vide, puisque la partition en une seule classe est galoisienne ; $\mathcal{C}$ est évidemment plus fine que la partition résultant du "produit" des partitions galoisiennes plus grossières qu'elle ; et cette partition "produit" est la classification galoisienne recherchée, puisque l'intersection de parties fermées est fermée.

La classification ξ_o s'obtient en considérant pour $\mathcal{C}$, la classification triviale dont chaque classe est composée d'un seul élément.

CONSTRUCTION

On cherche à construire la plus fine des classifications galoisiennes plus grossières qu'une partition donnée $\mathcal{C}$.

Notons C_x la classe de $\mathcal{C}$ contenant l'élément x de X, et P_x^i, la partie construite à l'étape i de l'algorithme et contenant x ; posons : $P_x^o = C_x$.

Algorithme :

- Si $i = 2k + 1$, $\quad P_x^i = \beta\alpha P_x^{i-1}$

Pour cela, il suffit de remarquer que :

$$x \in \beta\alpha P_{x'}^{i-1} \iff \alpha x \supset \alpha P_{x'}^{i-1}$$

- Si $i = 2k$, on construit la classification (P_x^i) *la plus fine* telle que :

$$P_x^{i-1} \cap P_{x'}^{i-1} \neq \emptyset \longrightarrow P_x^i = P_{x'}^i \ ;$$

en fait, si $P_x^{i-1} \cap P_{x'}^{i-1} \neq \emptyset$, on agrège ces deux parties ; puis, si $(P_x^{i-1} \cup P_{x'}^{i-1}) \cap P_{x''}^{i-1} \neq \emptyset$ on agrège les trois parties, et ainsi de suite.

Critères d'arrêt : Si on rencontre une partie P_x^i telle que $\alpha P_x^i = \emptyset$, on sait que la classification cherchée se compose de la seule classe X, puisqu'alors $\beta\alpha P_x^i = X$. Sinon, on arrête l'algorithme lorsque, pour tout x : $P_x^i = P_x^{i-1}$; alors la famille (P_x^i) de parties de X est la classification galoisienne cherchée, ce qu'il convient de démontrer.

Justification

A chaque étape i, on obtient une famille (P_x^i) de parties de X ; si i = 2k+1, c'est un recouvrement de fermés ; si i = 2k, c'est une partition.

Supposons que (P_x^i) soit une partition galoisienne :
- si i+1 = 2k+1 : $P_x^{i+1} = \beta\alpha P_x^i = P_x^i$, puisqu'on suppose tous les P_x^i fermés ;
- si i+1 = 2k : $P_x^i \cap P_{x'}^i \neq \emptyset$ ne se produit que si $P_x^i = P_{x'}^i$, puisqu'on suppose que (P_x^i) est une partition ; l'algorithme n'agrège donc aucune classe et $P_x^{i+1} = P_x^i$.

Supposons que, pour tout x, on ait $P_x^{i+1} = P_x^i$:
- si i+1 = 2k+1, (P_x^i) est une partition ; montrons qu'elle est galoisienne. Par construction $P_x^{i+1} = \beta\alpha P_x^i$; mais, par hypothèse, $P_x^{i+1} = P_x^i$; donc $P_x^i = \beta\alpha P_x^i$ est fermé.
- si i+1 = 2k, tout P_x^i est fermé ; si $P_x^i \cap P_{x'}^i \neq \emptyset$, alors $P_x^{i+1} = P_{x'}^{i+1}$, et par notre hypothèse, $P_x^i = P_{x'}^i$; (P_x^i) est donc une partition galoisienne.

Ainsi, l'algorithme s'arrête toujours sur une partition galoisienne. Reste à montrer que c'est la plus fine des partitions galoisiennes plus grossières que la classification $\mathcal{C}$ de départ.

Soit $\mathcal{G}$ la classification galoisienne associée à $\mathcal{C}$; notons G_x la classe où figure x ; il suffit de montrer que pour tout i, $P_x^i \subset G_x$:
- $P_x^0 = C_x \subset G_x$, par hypothèse ;
- si i = 2k+1, on a $P_x^i = \beta\alpha P_x^{i-1}$, mais si $P_x^{i-1} \subset G_x$, on a aussi $\beta\alpha P_x^{i-1} \subset \beta\alpha G_x = G_x$, par l'isotonie de $\beta\alpha$ et l'hypothèse que G_x est fermé.
- si i = 2k et, pour tout x : $P_x^{i-1} \subset G_x$, on a

$P_x^{i-1} \cap P_{x'}^{i-1} \subset G_x \cap G_{x'}$; si $P_x^{i-1} \cap P_{x'}^{i-1} \neq \emptyset$, alors $G_x \cap G_{x'} \neq \emptyset$ et, puisque $\mathcal{G}$ est une partition, $G_x = G_{x'}$; donc $P_x^{i-1} \cup P_{x'}^{i-1} \subset G_x = G_{x'}$. Par suite, P_x^i est obtenu par agrégation de fermés $P_{x'}^{i-1}$ tous contenus dans G_x : $P_x^i \subset G_x$.

Remarque : Pour trouver la classification galoisienne la plus fine, $\mathcal{G}_0$, il suffit de démarrer l'algorithme avec la classification $\mathcal{C}$ la plus fine, c'est à dire de poser :
$P_x^0 = C_x = \{x\}$.

EXEMPLE

Le tableau suivant correspond à une expérience où 14 sujets étaient invités à construire une phrase en utilisant certains termes choisis dans une liste. Cette liste constitue l'ensemble X ; Y est l'ensemble des sujets. La classification cherchée porte sur les mots.

X \ Y	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1 - Salaire	1	1	1	1	1	1	0	1	0	1	0	0	0	0
2 - Force de travail	0	1	0	0	0	0	0	0	0	1	1	0	0	1
3 - Profit	0	0	0	0	1	0	1	0	1	1	1	0	1	1
4 - Superflu	0	0	0	0	0	1	1	1	1	0	0	0	0	0
5 - Besoins	1	1	0	0	1	1	1	1	0	0	0	0	0	0
6 - Capitalistes	0	0	0	0	1	0	1	0	1	1	1	1	1	1
7 - Acheter	1	1	0	1	0	0	0	1	0	0	0	0	0	0
8 - Lutter	0	0	1	0	0	0	0	0	0	1	0	1	1	1
9 - Travailleurs	1	1	1	1	1	1	0	0	1	1	1	1	1	1

Etapes $P^i / \alpha P^i$

i=1 1/ABCDEFHJ, 29/BJKN, 36/EGIJKMN, 4/FGHI, 5/ABEFGH
6/EGIJKLMN, 71/ABDH, 89/CJLMN, 9/ABCDEFIJKLMN

i=2 17/ABDH, 289/JN, 36/EGIJKMN, 4/FGHI, 5/ABEFGH

i=3 17/ABDH, 2389/JN, 36/EGIJKMN, 4/FGHI, 5/ABEFGH

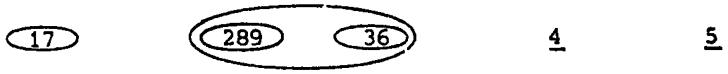
i=4 17/ABDH, 23689/JN, 4/FGHI, 5/ABEFGH

i=5 17/ABDH, 23689/JN, 4/FGHI, 5/ABEFGH

Ces étapes peuvent être illustrées comme suit (où chaque enveloppe réunit les éléments d'un fermé) :

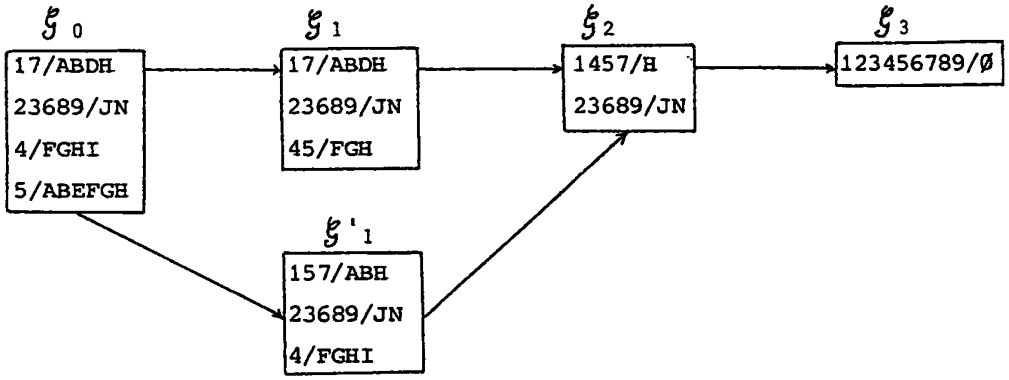
i=1 

i=2 17 / 289 / 36 / 4 / 5

i=3 

i=4 17 / 23689 / 4 / 5

La partition $\mathcal{P}_0$ ainsi obtenue est la plus fine qui réponde à la définition. On peut envisager d'étudier les partitions galoisiennes moins fines qui s'en déduisent. Soit $(G_i)_{i \in I}$, une partition galoisienne; s'il existe $J \subset I$ tel que $\beta\alpha(\bigcup_J G_j) = \bigcup_J G_j$, alors la partition composée de $(\bigcup_J G_j)$ et des classes G_i, $i \in I-J$, est évidemment galoisienne. On applique ce principe de proche en proche à partir de $\mathcal{P}_0$. Ainsi, dans notre exemple, on obtient :



APPROXIMATION

Parfois la classification $\mathcal{G}_o$ paraîtra insuffisamment fine. On sera donc tenté de chercher à éclater certaines classes. La structure permet d'envisager la prise de décisions qui ne soient pas "arbitrairement" fondées sur un calcul de distance. Ces décisions consistent à retirer certains éléments de X de façon à faire éclater certaines parties P^i. C'est notamment le cas d'éléments figurant dans plusieurs fermetures d'ensembles P^{i-1}, tel l'élément 9 dans notre exemple, qui joue un rôle déterminant dans l'obtention de la classe 23689.

Supposons qu'on prenne la décision de le retirer ; on obtient alors (en deux étapes) pour $\mathcal{G}_o$ le résultat suivant :

17/ABDH, 2/BJKN, 4/FGHI, 5/ABEFGH, 36/EGIJKMN, 8/CJLMN.

Bien entendu $\mathcal{G}_o$ est la classification galoisienne la plus fine pour cet ensemble et rien n'interdit de considérer comme précédemment les partitions moins fines dont la liste figure ci-dessous :

157/2/4/36/8	1457/2/36/8	17/2/36/45/8	17/2/4/356/8
157/2/346/8	1457/236/8	17/236/45/8	17/2/3456/8
157/236/4/8	1457/2/368	17/2/368/45	17/236/4/5/8
157/2/368/4	1457/2368	17/2368/45	17/2/368/4/5
157/2368/4	1257/36/4/8	17/2/346/5/8	17/2368/4/5
	1257/346/8		
	1257/368/4		

Ainsi l'observation des différentes étapes de l'algorithme donne le moyen de rechercher par approximations successives une classification faisant sens pour l'analyste.

Références citées

BARBUT, M. et MONJARDET, B., *Ordre et classification : algèbre et combinatoire*, Paris, Hachette, 1970, (deux tomes).

BIRKHOFF, G., *Lattice theory*, Colloquium publications, vol. XXV, AMS, Providence, 1940 (troisième édition, 1967).

Statistique et Analyse des Données
1 - 1978

ETUDE DE L'ALGORITHME DE WILKINSON EN ANALYSE DE VARIANCE

Astier Roger
I.U.T. d'Orsay - Département Informatique - 91406 ORSAY

RESUME

Présentation d'un algorithme d'analyse de variance traitant aussi bien les plans orthogonaux que ceux non orthogonaux. Ecriture de cet algortihme en Fortran.

1 - INTRODUCTION

Cet article reprend une méthode d'analyse de variance publiée par WILKINSON (1970) et JAMES et WILKINSON (1970) ; cette technique, peu connue en France, met en oeuvre un certain nombre d'idées qu'il nous a paru intéressant de résumer, d'autant plus, que c'est l'algorithme qui est employé dans la directive "ANOVA" de GENSTAT (NELDER, 1975). Enfin un travail de programmation, aboutissant à un sous-programme FORTRAN, a été effectué ; celui-là est disponible auprès de l'auteur, ou au laboratoire de Biométrie de l'INRA (78350 JOUY-EN-JOSAS).

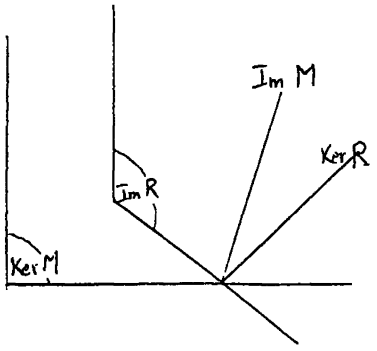
L'algorithme de WILKINSON est basé sur les deux idées suivantes :

1. une démarche itérative, dans laquelle les facteurs d'un modèle sont introduits progressivement, qui prend donc en compte la simplicité de la matrice d'incidence associée à un facteur en analyse de variance.

2. une inverse généralisée de matrice devant être trouvée, on détermine le polynome minimal associé à cette matrice pour y arriver.

Nous exposerons quelques considérations géométriques utilisées pour étudier la modification d'un modèle par introduction d'un facteur "simple" ; un bref aperçu du déroulement de l'algorithme est ensuite exposé (pour plus de détails se référer aux articles précédemment cités).
Un exemple est ensuite traité.

2 - CONSIDERATIONS GEOMETRIQUES



Dans R^n muni d'un produit scalaire, considérons deux projecteurs orthogonaux M et R, dont les images respectives sont notées Im M et Im R, et les noyaux respectifs Ker M et Ker R. Si x et y sont éléments de R^n $\langle x,y \rangle$ est leur produit scalaire ; si A et B sont deux sous espaces orthogonaux A $\oplus$ B représente leur somme directe (orthogonale) ; A' + B' représente la somme ou la somme directe de deux sous espaces de R^n.

2.1 - Diagonalisation de MRM.

Proposition

F étant le supplémentaire orthogonal dans Im M de (Ker R ∩ Im M), e_1, ... e_k étant les valeurs propres non nulles (si elles existent) de MRM, $F_{e_1} \ldots F_{e_k}$ les sous espaces propres qui leur sont respectivement associés :

i) $F \neq \{0\} \Leftrightarrow$ MRM admet une valeur propre nulle.

ii) $F = F_{e_1} \oplus \ldots \oplus F_{e_k}$ et $\text{im } M = F \oplus (\text{Ker } R \cap \text{Im } M)$ sont des sommes directes orthogonales.

iii) $e_1, \ldots, e_k$ sont des réels positifs ; de plus si $\theta_1 \ldots \theta_k$ sont les angles canoniques, distincts de $\frac{\pi}{2}$, entre Im R et Im M, alors $e_i = \cos^2\theta_i$ (pour i = 1...k).

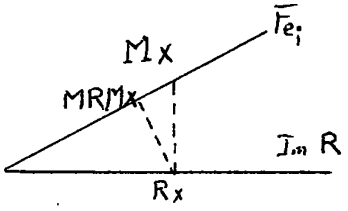
iiii) le sous espace propre de MRM associé à la valeur propre 0 est $\text{Ker } M \oplus (\text{Ker } R \cap \text{Im } M)$.

Démonstration

MRM est un endomorphisme symétrique, donc diagonalisable, avec des sous espaces propres deux à deux orthogonaux. Si N est la restriction de MRM à Im M, N est aussi un endomorphisme symétrique. N et M ont les mêmes sous espaces propres associés à des valeurs propres non nulles ($\frac{1}{e}$ MRMx = x $\overset{\leftarrow}{\rightarrow}$ (x ∈ Im M et $\frac{1}{e}$ Nx = x). Cherchons le sous espace propre de N associé à la valeur propre de 0. Ce sous espace contient Ker R ∩ Im M ; si maintenant x est vecteur propre de N associé à la valeur propre 0 alors : Mx = x car x ∈ Im M ; MRMx = 0 ;
calculons <Rx, Rx> = <Rx, x> = <RMx, Mx> = <MRMx, x> = 0 : donc Rx = 0 et x ∈ Ker R ; le sous espace propre de N associé à la valeur porpre 0 est (Ker ∩ Im M).
Si MRM admet 0 comme seule valeur propre, alors il en est de même pour N donc Ker R ∩ Im M = Im M, donc F = {0} et réciproquement.

Le point ii est démontré en utilisant la restriction N, endomorphisme symétrique.

Les sous espaces propres F_{e_1}, ..., F_{e_k} sont les sous espaces canoniques entre Im R et Im M. Si x est un élément de F_{e_i} de norme 1, alors :



e_i = <MRMx, x> = <RMx, Mx> = <Rx, x> = $\cos^2\theta$ i
Ainsi le point iii est démontré.
Le point iiii se déduit du fait que l'endomorphisme MRM est diagonalisable.

2.2 - Diagonalisation de MR.

Proposition

i) MR et MRM ont les mêmes valeurs propres non nulles et les mêmes sous espaces propres respectivement associés.

ii) MR est diagonalisable et a les mêmes valeurs propres que MRM ;
R^n = Ker R + (Im R ∩ Ker M) + F est une somme directe (pas nécessairement orthogonale).

Démonstration

i est évident en écrivant pour x vecteur propre associé à la valeur propre non nulle e :

$$\frac{1}{e} MRx = \overset{\leftarrow}{\rightarrow} x \in \text{Im } M \text{ donc } \frac{1}{e} MRMx = x$$

$$\frac{1}{e} MRMx = \overset{\leftarrow}{\rightarrow} x \in \text{Im } M \text{ donc } \frac{1}{e} MRx = x$$

On peut aborder ii en disant que Ker R $\oplus$ (Im R $\cap$ Ker M) est inclus dans le sous espace propre de MR associé à la valeur propre 0 ; montrons de plus que les sous espaces :

$$F_{e_i}, \ldots, F_{e_k}, \text{Ker } R \oplus (\text{Im } R \cap \text{Ker } M)$$

forment une somme directe de R^n. D'une part ces sous espaces sont en somme directe car ils sont inclus dans des sous espaces propres associés à des valeurs propres distinctes de l'endomorphisme MR.

D'autre part, calculons la somme des dimensions de chacun d'eux (en utilisant 2.1-II et le fait que l'orthogonal de (Im R $\cap$ Ker M) est $(\text{Im } R)^{\perp} + (\text{Im } M)^{\perp}$ soit (Ker R + Im M) :

$$\begin{aligned}
&\dim (F_{e_1}) + \ldots + \dim (F_{e_k}) + \dim \left[\text{Ker } R \oplus (\text{Im } R \cap \text{Ker } M)\right] \\
&= \dim (F) + \dim (\text{Ker } R) + \dim (\text{Im } R \cap \text{Ker } M) \\
&= \left[\dim (\text{Im } M) - \dim (\text{Ker } R \cap \text{Im } M)\right] + \dim (\text{Ker } R) + \left[n - \dim(\text{Ker } R + \text{Im } M)\right] \\
&= n + \dim (\text{Im } M) + \dim (\text{Ker } R) - \left[\dim (\text{Ker } R + \text{Im } M) + \dim (\text{Ker } R \quad \text{Im } m)\right] \\
&= n
\end{aligned}$$

ii et iii sont alors démontrés.

2.3 - Polynome minimal de MRM ou de MR

Proposition

En supposant Im M non inclus dans Ker R et Im M distinct de R^n, le polynome minimal de MRM ou de MR est de degré au moins égal à 2 ; on peut l'écrire :

i) $m(x) \equiv x + \sum_{i=1}^{k} \alpha_i x^{i+1}$

ii) $m(x) \equiv x - xq(x)x$, et alors q(MR) est un inverse généralisé de MR.

iii) $m(x) \equiv xp(x)$, et alors $p(x) = 1 - xq(x) = (1 - \frac{1}{e_1} x) \ldots (1 - \frac{1}{e_k} x)$;

p(MR) est une projection sur Ker R $\oplus$ (Im R $\cap$ Ker M) parallèlement à F.

Démonstration

La condition (Im M non inclus dans Ker R) est équivalente à MRM $\neq$ 0 donc il existe au moins une valeur propre non nulle pour MRM ou pour MR ; la condition (Im M distinct de R^n) est équivalente à l'existence de 0 comme valeur propre de MRM.

Lorsque les deux conditions sont réalisées, il existe au moins deux valeurs propres distinctes, le polynome minimal est de degré au moins égal à 2. Ce polynome a nécessairement la

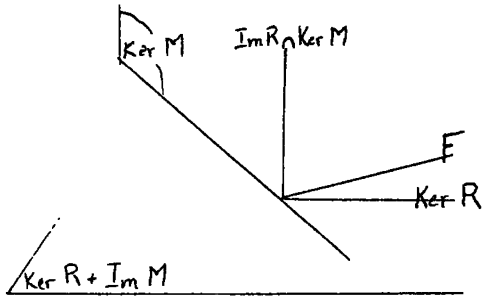
forme donnée en i (à une constante multiplicative non nulle près) car :

- 0 est une racine du polynome donc pas de terme constant ;
- 0 est une racine simple, comme toutes les racines du polynome minimal d'un endomorphisme diagonalisable ; donc le terme de plus bas degré est de degré 1.

En utilisant la forme ii on trouve l'égalité MR - MRq(MR) MR = 0 d'où l'inverse généralisé.

La forme iii dégage p(x) comme polynome ayantpour racines e_1, e_2, ... e_k ; pour t vecteur propre de MR associé à la valeur propre e on peut écrire p(MR)t = t = p(e)t; avec p(e) = 0 si e est une valeur propre non nulle et p(o) = 1. En utilisant les résultats sur la diagonalisation de MR on vérifie immédiatement que l'endomorphisme p(MR) est nul sur F, et se restreint à l'identité sur Ker R ⊕ (Im R ∩ Ker M).

Remarque : On obtient facilement les résultats suivants :



a) I-p(MR) = q(MR) MR est une projection sur F parallèlement à Ker R ⊕ (Im R Ker M)

b) Rp(MR) est une projection orthogonale sur Im R∩Ker M

b') I-Rp(MR) est une projection orthogonale sur Ker R + Im M

c) (I-R)p(MR) est une projection sur Ker R parallèlement à (Im R ∩ Ker M) + F

3 - MODIFICATION D'UN MODELE PAR INTRODUCTION D'UN FACTEUR

3.1 - Présentation

Soit le modèle : $\mathcal{M}_o$

$Y = X_o t_o + EPS$

Y , EPS v.a. à valeurs dans R^n ; E(EPS) = 0

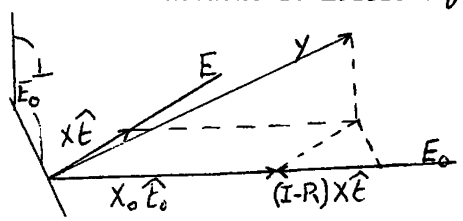
var EPS = $\sigma^2 W$;

R^n muni du produit scalaire W^{-1};

E_o sous espace engendré par les colonnes de X_o;

Supposons connue la projection orthogonale R sur $E_0^{\perp}$; I-R est la projection orthogonale sur E_o et nous savons estimer $X_o t_o$ dans ce modèle.

Considérons maintenant le modèle : $\mathcal{M}$



$Y = X_o t_o + Xt + EPS$

Y, EPS à valeurs dans R^n ;

E(EPS) = -O : var EPS = σ^2 W ;

E sous espace engendré par les colonnes de X ;

$X'W^{-1}X$matrice diagonale inversible.

Pour ce nouveau modèle nous devons déterminer la projection orthogonale sur $E + E_o$ permettant d'avoir une estimation, notée $X_o\hat{t}_o + X\hat{t}$, de $X_o t_o + Xt$; nous devons aussi décomposer cette estimation en un élément de E_o et un élément de E pour obtenir séparément $X_o\hat{t}_o$ et $X\hat{t}$.

Nous calculerons aussi la projection orhtogonale sur $(E_o + E)^{\perp}$ afin d'avoir une estimation de σ^2. Enfin nous calculerons les covariances des estimateurs de t_o et de t, dans ce modèle $\mathcal{M}$.

La projection orthogonale M sur E est aisément calculable: $(X(X'W^{-1}X)^{-1}X'W^{-1})$; on peut écrire :

$Y = X_o t_o + Xt + EPS = X_o \hat{t}_o + X\hat{t} + EPS$

$RY = RX\hat{t} + \widehat{EPS}$ car $R X_o \hat{t}_o = 0$ puisque $X_o \hat{t}_o$ est élément de E_o

$R\,\widehat{EPS} = \widehat{EPS}$ car $\widehat{EPS}$ à valeur dans $(E_o + E)^{\perp} = E_o^{\perp} \cap E^{\perp}$

$MRY = MRX\hat{t}$ car $\widehat{MPES} = 0$

Ainsi l'estimateur de Xt est trouvé à partir d'une inverse généralisée de MR, que l'on sait trouver à partir du polynome minimal de MR, d'après la paragraphe ci-dessus. Reportons nous à ce paragraphe.

3.2 - Utilisation des résultats précédents

La projection orthogonale R a pour image $E_o^{\perp}$ et pour noyau E_o, la projection orthogonale M a pour image E et pour noyay $E^{\perp}$.

Par application des résultats démontrés nous pouvons écrire :

$E = (E_o \cap E) \oplus F$ introduit F comme somme directe des sous espaces propres de MR associés aux valeurs propres non nulles. (voir 2.1. ii)

$R^n = E_o + (E_o^{\perp} \cap E^{\perp}) + F$ présente $E_o + (E_o^{\perp} \cap E^{\perp})$ comme sous espace propre de MR associé à la valeur propre 0. (voir 2.2. ii)

Cette dernière égalité montre, en outre, que $E_o + F$ a même dimension que l'orthogonal de $E_o^{\perp} \cap E^{\perp}$, soit $E_o + E$; comme de plus $E_o + F$ est inclus dans $E_o + E$ on en conclut que :

E_o et F sont en somme directe dans $E_o + E$

3.3 - Estimation dans le modèle $\mathcal{M}$

Nous supposons dans un premier temps que E n'est pas inclus dans E_o ; les résultats du paragraphe 2.3. faisant intervenir le polynome minimal de MR, et le polynome p sont alors :

- I-Rp(MR) est la projection orthogonale sur $E_o + E$ donc :

$$X_o \hat{t}_o + X\hat{t} = \left[I - Rp(MR \right] Y$$

- I- p(MR) est la projection sur F parallèlement à $E_o + (E_o + E)^{\perp}$, projection oblique en général donc :

$$X\hat{t} = \left[I - p(MR) \right] Y$$

- (I-R)p(MR) est la projection sur E_o parallèlement à $F + (E_o + E)^{\perp}$ projection oblique en général donc :

$$X_o \hat{t}_o = (I-R)p(MR)\ Y$$

On peut constater de plus, en utilisant l'égalité : $(I-R)\,p(MR) = I-R-(I-R)\,(I-p(MR))$ que

$$X_o \hat{t}_o = (I-R)\,Y - (I-R)\ X\hat{t}$$

Ainsi l'estimation du facteur t_o dans le modèle de $\mathcal{M}$ est la somme de l'estimation de ce facteur dans le modèle $\mathcal{M}_o$ et d'un terme "correctif " tenant compte de l'existence du facteur t.

- Rp(MR) est la projection orthogonale sur $(E_o + E)^{\perp}$ donc :
  $Y'W^{-1}$ Rp(MR) Y a pour espérance σ^2 trace (Rp(MR)).

Dans le cas où E est inclus dans E_o, alors Im M est inclus dans Ker R et le sous espace F est réduit à { 0 }. L'estimation $\hat{Xt}$ est alors nulle. En fait, l'introduction du facteur t n'apporte rien sur la connaïssance de EY ; on dit que le facteur t est totalement confondu avec le facteur t_o.

Remarques

1) le sous espace $E_o + E$ est donc décomposé en une somme directe de E_o (contenant $\hat{Xt}_o$) et de F, orthogonal à $E_o \cap E$ dans E ; l'estimation $\hat{Xt}$ se fait dans F ; les contraintes permettant l'estimation se traduisent par l'orthognalité de $\hat{Xt}$ avec $E_o \cap E$.

2) l'ordre "d'introduction" des facteurs dans le modèle joue un rôle important ; si nous écrivons $E_o = F_o \oplus (E_o \cap E)$ (somme directe orthogonale) alors :

- $E_o + E = F_o + (E_o \cap E) + F$, le second membre étant une somme directe.
- Si $X_o\hat{t}_o + \hat{Xt} = a + b + c$ où $a \in F_o$, $b \in E_o \cap E$, $c \in F$ alors :

dans le modèle "$Y = X_o t_o + Xt + EPS$" on estime $X_o t_o$ par a+b on estime Xt par c;
dans le modèle "$Y = Xt + X_o t_o + EPS$" on estime $X_o t_o$ par a, on estime Xt par b+c.

3) Examinons le cas où $t_o = \left(\begin{smallmatrix} t\alpha \\ t\beta \end{smallmatrix}\right)$ représente deux facteurs α et β et t représente l'intéraction entre ces deux facteurs ; alors E_o est inclus dans E et le sous espace dans lequel est estimée l'intéraction est ici orthogonal à E_o (qui est aussi $E_o \cap E$) ; nous trouvons une estimation de l'interaction orthogonale aux estimations des facteurs principaux α et β ; ceux sont les contraintes habituelles imposées à l'interaction.

4) Si E_o et F sont othogonaux alors F est inclus dans Im R et il existe un seul angle (o mod (π)) différent de $\frac{\pi}{2}$ entre Im M et Im R. Ainsi MR a une seule valeur propre non nulle qui vaut 1, le sous espace propre associé étant Im R $\cap$ Im M $\equiv E_o^{\perp} \cap E$.
Le polynome minimal est MR(I-MR) = 0 ; $F = E_o^{\perp} \cap E$ et donc $(I-R)\hat{Xt} = 0$
les estimations de t_o dans les modèles $\mathcal{M}_o$ et $\mathcal{M}$ sont les mêmes.

3.4 - Modifications des sommes de carrés

En notant $\| x \|^2$ la norme d'un élément de R^n pour le produit scalaire W^{-1}, nous pouvons écrire :

$\| (I-R)\ Y \|^2$ est la contribution des facteurs à la somme des carrés dans le modèle $\mathcal{M}$.

$\| (I-Rp(MR))\ Y \|^2$ est la contribution des facteurs à la somme des carrès dans le modèle $\mathcal{M}$.

Comme $\left[I-Rp(MR) \right] . Y = X_o\hat{t}_o + \hat{Xt}$
$= (I-R)\ Y - (I-R)\ \hat{XT} + X_2\hat{t}_2$
$= (I-R)\ Y + R\ X\ \hat{t}$

et que les vecteurs (I-R) Y et R X $\hat{t}$ sont orthogonaux, on peut écrire :

$$\| \left[I-Rp(MR) \right] Y \|^2 = \| (I-R)\ Y \|^2 + \| R\ X\ t \|^2.$$

Ainsi $\|R \times \hat{t}\|^2$ apparaît comme contribution du facteur t (introduit après t_o) à l'accroissement de la somme des carrés expliquée par les facteurs.

3.5 - Covariance des estimateurs

En supposant var (EPS) $=\sigma^2 I$ pour simplifier les calculs, on trouve :

i) var $(X\hat{t}) = \sigma^2 V$ où $V = q(MR)M + p'(o)p(MR)M$

ii) cov $(X_o\hat{t}_o, X\hat{t}) = \sigma^2 (I-R) V$

iii) var $(X_o\hat{t}_o) = \sigma^2(I-R) + \sigma^2(I-R) V (I-R)$

Démonstration

$X\hat{t} = \left[I - p(MR)\right] Y$ et var $Y = \sigma^2 I$ donc :

var $(X\hat{t}) = \left[I-p(MR)\right]$ var $Y \left[I-p(MR)\right]' = \sigma^2\left[I-p(MR)\right]\left[I-p(RM)\right]$

ou var $(X\hat{t}) = \sigma^2\left[I-p(MR)\right] RMq(RM) = \sigma^2\left[I-p(MR)\right]Mq(RM) = \sigma^2\left[I-p(MR)\right] q(MR)M$

or $p(MR)q(MR)M = p(MR)\left[-\alpha_1 M - \sum_{i=2}^{k} \alpha_i (MR)^{i-1} M\right]$ avec les notations du paragraphe 2.3.

$= -\alpha_1 p(MR)\, M - p(MR)\, MR \sum_{i\ 2} {}_i(MR)^{i-2}M$

$= -\alpha_1 p(MR)\, M$ car $p(MR)\, MR = 0$

de plus $\alpha_1 = p'(o)$ d'où le résultat i)

Les résultats ii) et iii) peuvent se démontrer de façon analogue.

Etudions la quantité $\gamma' V \gamma$ quand γ est un élément particulier de Im M, c'est à dire de E :

- si $\gamma \in E_o \cap E$: $V\gamma = q(MR)\,\gamma + p'(o)\, p(MR)\,\gamma$
  $= \left[q(o) + p'(o)\, p(o)\right]\gamma$
  $= 0$ car $p(o) = I$ et $q(o) = -p'(o)$
  donc $\gamma' V \gamma = 0$

- si $\gamma \in F_{e_i}$: $V\gamma = q(MR)\,\gamma + p'(o)\, p(MR)\gamma$
  $= q(e_i)\,\gamma + p'(o)\, p(e_i)\gamma$
  $= \frac{1}{e_i}\gamma$ car $p(e_i) = 0$ et $q(e_i) = \frac{1}{e_i}$
  donc $\gamma' V\gamma = \frac{1}{e_i}\gamma'\gamma$

La variance d'une forme linéaire estimable $X\hat{t}$, (γ est élément de F), normalisée par $\gamma'\gamma = I$ est donc $\left[\frac{1}{e_1}\,\gamma_1'\,\gamma_1 + \frac{1}{e_2}\,\gamma_2'\,\gamma_2 + \ldots + \frac{1}{e_k}\,\gamma_k'\,\gamma_k\right]\sigma^2$

si $\gamma = \gamma_1 + \ldots + \gamma_k$ correspond à la décomposition $F = F_{e_1} \oplus \ldots\ldots \oplus F_{e_k}$.

Ainsi var $(\gamma' X\hat{t})$ avec $\gamma'\gamma = I$ est extrêma minimum si γ appartient à un sous espace propre F_{e_i}, et vaut alors $\frac{1}{e_i}\sigma^2$; e_i <u>est un facteur d'efficacité dans le modèle</u> $\mathcal{M}$, du facteur t.

4 - DEROULEMENT DE L' ALGORITHME

Envisageons le modèle, ayant f facteurs fixes.

$$Y = X_1 t_1 + \ldots + X_f t_f + EPS$$

où Y, EPS sont des variables aléatoires à la valeurs dans R^n;

$t_1, t_2 \ldots t_f$ sont, respectivement, éléments de $R^{p1}, R^{p2}, \ldots R^{pf}$;

$E(EPS) = 0$; $var(EPS) = \sigma^2 W$ avec W inversible ;

$X'_1 W^{-1} X_1, X'_2 W^{-1} X_2, \ldots, X'_f W^{-1} X_f$ sont inversibles.

Wilkinson suggère, pour l'étude de ce modèle, d'effectuer d'abord un "analyse muette" consistant à étudier (f-1) modèles, obtenus par adjonction progressive d'un seul facteur par rapport au modèle précédent. Il faut donc déterminer (f-1) polynomes minimaux : on l'utilise pour celà :

Proposition

Soient A une matrice de format (p , p) diagonalisable et à coefficients rationnels, $e_1, e_2, \ldots e_k$ ses valeurs propres ($i \neq j \Leftrightarrow e_i \neq e_j$), t un élément de R^p dont les coordonnées sont les puissances successives d'un nombre transcendant (par exemple $t' = (1, \pi, \ldots, \pi^{p-1})$. Alors :

i) $t, At, \ldots, A^{k-1}t$ sont indépendants.

ii) $A^k t - \sum_{i=0}^{k-1} \lambda_i A^i t = 0$ et le polynome minimal de A est $x^k - \sum_{i=0}^{k-1} \lambda_i x^i$

Démonstration

Les valeurs propres $e_1, e_2, \ldots, e_k$ sont algébriques puisque A est à coefficients rationnels.

Les sous espaces propres de A formant une somme directe de R^p (A est diagonalisable), je peux écrire :

$$t = a_1 + a_2 + \ldots + a_k \text{ où } Aa_i = e_i a_i (i = 1 \ldots k).$$

Alors les a_i sont nuls. Supposons en effet avoir $a_1 = 0$; alors cela entraine ,

$$\prod_{j=2}^{k} (A - e_j I)\, t = 0$$

Donc $\prod_{j=2}^{k} (A - e_j I) = 0$ car sinon l'égalité ci-dessus donnerait un polynome en π, nul avec des coefficients algébriques non tous nuls. Or $\prod_{j=2}^{k} (A - e_j I)$ est non nul puisque e_1 est une valeur propre de A.

On peut donc dire que E, sous espace engendré par $a_1, a_2, \ldots a_k$ admet pour base $B = (a_1, a_2, a_k)$, puisque ces vecteurs sont non nuls, et appartiennent respectivement à des sous espaces en somme directe. On peut calculer $A^i t = (e_1)^i a_1 + \ldots + (e_k)^i a_k$. Les éléments $A^i t$ appartiennent donc à E. $t, A_1 t, \ldots A^{k-1}$ sont indépendants, puisque le déterminant de ces vecteurs dans la base B est un déterminant de Vandermonde, et forment une base de E. $A^k t$ est donc combinaison linéaire de $t, At \ldots A^{k-1} t$:

$$A^k t = \lambda_o t + \lambda_1 At + \ldots + \lambda_{k-1} A^{k-1} t$$

$$\pm \begin{pmatrix} (e_1)^k \\ \\ (e_k)^k \end{pmatrix} \begin{vmatrix} 1 & 1 & \ldots & 1 \\ e_1 & e_2 & \ldots & e_k \\ (e_1)^{k-1} & (e_2)^{k-1} & \ldots & (e_k)^{k-1} \end{vmatrix} \begin{pmatrix} \lambda_o \\ \lambda_1 \\ \lambda_{k-1} \end{pmatrix}$$

Les coefficients $\lambda_o, \lambda_1 \ldots \lambda_{k-1}$ sont algébriques, car solutions d'un système à coefficients algébriques, et l'égalité $(A^k - \sum_{i=0}^{k-1} \lambda_i A^i)\, t = 0$ avec les coordonnées de t transcendantes, entraine :

$$A^k - \sum_{i=0}^{k-1} \lambda_i A^i = 0$$

Ainsi le polynome minimal est obtenu en cherchant la relation de dépendance entre $t, At, \ldots, A^k t$ (par procédé d'orthogonalisation) faisant intervenir le plus petit k.

L'algorithme se poursuit par la détermination, pour chaque facteur fixe, du degré de liberté de son estimateur, ainsi que de la matrice de covariance générale de ces estimateurs (à σ^2 près).

Enfin on utilise les résultats de l'"analyse muette" pour obtenir, à partir d'une variable expérimentale, les estimations des effets des facteurs fixes, ainsi qu'une décomposition en somme de carrès.

5 - EXEMPLES

Nous étudions un exemple, dans lequel apparaissent facilement les sous espaces $E_o \cap E$ et F en particulier. Nous intervertissons ensuite l'ordre d'apparition des facteurs.

Considérons donc le cas de deux facteurs à trois niveaux, t_o et t, dont les matrices d'incidence, respectivement X_o et X sont données ci-dessous.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} \quad ; \quad X_o = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad ; \quad X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

base de $E_o \cap E$: $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\}$;

$$\text{BASE DE F} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ -3 \end{pmatrix} \right\}$$

Les sous espaces E_o et E ont intersection de dimension 2, dont on peut trouver facilement une base ; F orthogonal de $E_o \cap E$ dans E est donc de dimension 1 ; on détermine un vecteur de E non nul, orthogonal à $E_o \cap E$, pour avoir une base de F (produit scalaire euclidien de R^n). Les projections orthogonales, (I-R) sur E_o, R sur E_o et M sur E se calculent facilement.

$$(I-R)Y = \frac{1}{2}\begin{bmatrix} y_1 + y_2 \\ y_1 + y_2 \\ y_3 + y_4 \\ y_3 + y_4 \\ y_5 + y_6 \\ y_5 + y_6 \end{bmatrix} \quad ; \quad RY = \frac{1}{2}\begin{bmatrix} y_1 - y_2 \\ -y_1 + y_2 \\ y_3 - y_4 \\ -y_3 + y_4 \\ y_5 - y_6 \\ -y_5 + y_6 \end{bmatrix} \quad ; \quad MY = \frac{1}{6}\begin{bmatrix} 3y_1 + 3y_2 \\ 3y_1 + 3y_2 \\ 2y_3 + 2y_4 + 2y_5 \\ 2y_3 + 2y_4 + 2y_5 \\ 2y_3 + 2y_4 + 2y_5 \\ 6y_6 \end{bmatrix} ;$$

$$MRY = \frac{1}{6}\begin{bmatrix} 0 \\ 0 \\ y_5 - y_6 \\ y_5 - y_6 \\ y_5 - y_6 \\ -3y_5 + 3y_6 \end{bmatrix} ; \quad (MR)^2 Y = \frac{1}{9}\begin{bmatrix} 0 \\ 0 \\ y_5 - y_6 \\ y_5 - y_6 \\ y_5 - y_6 \\ -3y_5 + 3y_6 \end{bmatrix}$$

Dans le cas du modèle $\mathcal{M}$: $Y = X_o t_o + Xt + EPS$ (avec $E(EPS) = 0$; $var(EPS) = \sigma^2 I$), on obtient l'estimateur $\hat{Xt}$ dans le sous espace F, de dimension 1 ; pour ce modèle cet estimateur a pour degré de liberté 1. Nous le calculons précisemment à partir du polynome minimal de MR. Les calculs de MRY et $(MR)^2 Y$ font apparaitre :

- polynome minimal $MR - \frac{3}{2}(MR)^2$;
- $p(MR) = I - \frac{3}{2}(MR)$; $q(MR) = \frac{3}{2} I$

Les estimations sont obtenues par $\hat{Xt} = \frac{3}{2} MRY$ et $\hat{Xt}_o = (I-R)Y - (I-R)\hat{Xt}$

$$\hat{Xt} = \frac{1}{4}\begin{bmatrix} 0 \\ 0 \\ y_5 - y_6 \\ y_5 - y_6 \\ y_5 - y_6 \\ -3y_5 + 3y_6 \end{bmatrix} ; \quad (I-R)\hat{Xt} = \frac{1}{4}\begin{bmatrix} 0 \\ 0 \\ y_5 - y_6 \\ y_5 - y_6 \\ -y_5 + y_6 \\ y_5 + y_6 \end{bmatrix} ; \quad X_o\hat{t}_o = \frac{1}{4}\begin{bmatrix} 2y_1 + 2y_2 \\ 2y_1 + 2y_2 \\ 2y_3 + 2y_4 - y_5 + y_6 \\ 2y_3 + 2y_4 - y_5 + y_6 \\ 3y_5 + y_6 \\ 3y_5 + y_6 \end{bmatrix}$$

Nous pouvons, de la même façon que précédemment, étudier le modèle :

$$\mathcal{M}' : Y = Xt + X_o t_o + EPS$$

Les estimations, dans ce cas, notées $X_o\tilde{t}_o$ et $\tilde{Xt}$ sont alors :

$$\tilde{Xt} = \frac{1}{4}\begin{bmatrix} 2y_1 + 2y_2 \\ 2y_1 + 2y_2 \\ y_3 + y_4 + 2y_5 \\ y_3 + y_4 + 2y_5 \\ y_3 + y_4 + 2y_5 \\ y_3 + y_4 - 2y_5 + 4y_6 \end{bmatrix} \qquad X_o\tilde{t}_o = \frac{1}{4}\begin{bmatrix} 0 \\ 0 \\ y_3 + y_4 - 2y_5 \\ y_3 + y_4 - 2y_5 \\ -y_3 - y_4 + 2y_5 \\ -y_3 - y_4 + 2y_5 \end{bmatrix}$$

$$X_o\hat{t}_o - X_o\tilde{t}_o = \frac{1}{4}\begin{Bmatrix} 2y_1 + 2y_2 \\ 2y_1 + 2y_2 \\ y_3 + y_4 + y_5 + y_6 \\ y_3 + y_4 + y_5 + y_6 \\ y_3 + y_4 + y_5 + y_6 \\ y_3 + y_4 + y_5 + y_6 \end{Bmatrix}$$

On peut vérifier que $X_o\hat{t}_o + X\hat{t} = X_o\tilde{t}_o + X\tilde{t}$ est la projection orthogonale sur $E_o + E$; de plus $X_o\hat{t}_o - X_o\tilde{t}_o$ est élément de $E_o \cap E$, projection orthogonale sur $E_o \cap E$ de Y.

7 - BIBLIOGRAPHIE

(1) JAMES, WILKINSON (1971). Factorisation of the residueal operator, and canonical decomposition of non-orthogonal factors in the Analysis of Variance. Biometrika, 58, 2 p. 279

(2) NELDER (1975). Genstat Reference Manual - Statistics Department Rothamsted Experimental Station - Harmenden Herfordshire - ENGLAND

(3) ROGERS WILKINSON (1974). Analysis of designed experiments Genstat User's Guide, 2.

(4) WILKINSON (1970). A general recursive procedure for analysis of variance. Biométrika, 57, 1p. 19.